TR-IT-0234

HMnet 作成における種々の音響分析の比較

Comparison of Different Types of Preprocessing for HMnet Generation

柘植 覚 Satoru Tsuge
シンガー・ハラルド Harald Singer
深田 俊明 Toshiaki Fukada

1997 年 9 月 15 日

音声認識システムにおける特徴パラメータの選択は、認識性能を左右する重要な問題である。そのため、本報告において種々ある特徴パラメータの中から、現在広く使用されている LPC(メル) ケプストラム、MFCC を選択し認識性能の比較を行った。認識性能の比較はサンプル周波数 12kHz、16kHz で行い、12kHz では MFCC、16kHz では LPC メルケプストラムが次数の変動に関わらず安定した認識性能を示した。また、ベースライン実験 (TR-IT-0206 の音響モデル) において無音部分の誤りが数多く出現していた。そのため、ポーズモデルの学習条件、状態数の変更を行った。連結学習を行わず、複数状態にした結果、音素認識率が 2.9(%) 向上し、74.14(%) となった。さらに、HMnet 初期状態の変動に対する認識性能の比較を行った。当該音素の共有を認めた初期経路長 3、27 状態を初期状態とし、最大経路長を 4 とした HMnet が最も高い認識性能を示した。初期状態による認識性能の変化は少なく、むしろ初期経路長、最大経路長に認識性能の影響が大きいことが分かった。

目次

1 はじめに

音声認識システムにおける特徴パラメータの選択は、認識性能を左右する重要な問題である。本報告では、種々ある特徴パラメータの中から、現在広く用いられているLPC(メル)ケプストラム、MFCCを用い、雑音のないクリーンな環境で収録された自然発話データベースに対して、認識性能の比較を行った。また、無音部分を表すポーズモデルの変更による認識性能の比較を行った。

以下、第2章では、ベースライン実験[4]について述べ、第3章では、特徴パラメータ変更による認識性能の比較ついて述べ、第4章では、ポーズモデル変更による認識性能の比較について述べ、最後に第5章では、性能比較実験のまとめを述べた。

その他、パワー項、HMnet作成条件、HMnet作成時学習状態の変化などの検討も行ったが、ベースライン実験の性能を越えることができなかった。これらの実験内容・結果については付録として示しておく。

2 ベースライン実験

本報告は1997年3月に提出されたテクニカルレポートTR-IT-0206 [4]に示されている実験をベースラインとし、種々の比較実験を行った。実験方法は同テクニカルレポートに準じているため、実験方法、実験に用いられたアルゴリズム等の詳細は省いた。本報告に必要な条件を簡単に以下に述べる。

Travel Arrangementをタクスとする自然発話音声データベース[2]から、学習データに、230名(男性100名、女性130名)、総音素数約12万(約200分)を用い、評価データに、学習に用いられていない42名(男性17名, 女性25名)、総音素数約2万(約40分)を用いた。表1に示した条件で、ML-SSSアルゴリズム[3]を用い、初期状態数27、初期経路長3から総状態数800、各5混合、最大経路長4のHMnet(音素環境依存HMM)の有音モデルを作成し、これに1状態10混合の無音モデルを付加したものを音響モデルとした。この音響モデルを用い音素認識実験を行った。結果を表2に示す。また、各話者毎の結果、特徴パラメータ作成、認識に用いたコンフィグレーションを付録Dに添付する。

後述の比較実験には、ベースライン実験と同じ学習データ、評価データを用いた。

表1: 音響分析

サンプル周波数	12kHz (downsampled from 16kHz)
プリエンファシス	0.98
量子化ビット数	16bit
フレームシフト	10 msec
フレーム長	20 msec
分析窓	Hamming 窓
音響ベクトル	16次LPCケプストラム, 対数パワー, 及びその一次回帰係数(計34次元)
Δ窓	triangular 100 msec (9 frames)[6]

表2: 基準実験結果

	男性	女性	全体
音素認識率	66.6(%)	74.0(%)	71.2(%)

3 音響分析に関する実験

3.1 LPC メルケプストラム

現在まで特徴パラメータとして LPC メルケプストラムを使用する場合、LPC 係数から LPC ケプストラムを計算した後、メル尺度に変換を行う再帰式を用い計算を行っていた。しかし、この計算方法では途中に用いる LPC ケプストラムの次数の変化が認識性能に影響を与えてしまう。そのため、直接 LPC 係数から LPC メルケプストラムを計算する、メルケプストラムの再帰的計算法 [5] を用い実験を行った。式 (1) ～ (3) に用いた再帰式を示す。この再帰式は、LPC ケプストラムの次数を無限大にしたのと等価であるため、打ち切り誤差による影響をなくすことができる。

$$\tilde{a}^{(i)}(m) = \begin{cases} a(-i) + \alpha\,\tilde{a}^{(i-1)}(0), & m = 0 \\ (1-\alpha^2)\,\tilde{a}^{(i-1)}(0) + \alpha\,\tilde{a}^{(i-1)}(1), & m = 1 \\ \tilde{a}^{(i-1)}(m-1) + \alpha\,(\tilde{a}^{(i-1)}(m) - \tilde{a}^{(i)}(m-1)), & m = 2,3,\ldots,N \end{cases}$$
$$i = -M,\ldots,-1,0 \qquad (1)$$

$$\tilde{K} = K/\tilde{a}^{(0)}(0), \qquad \tilde{a}(m) = \tilde{a}^{(0)}(m)/\tilde{a}^{(0)}(0), \quad 1 \le m \le N \qquad (2)$$

$$\tilde{c}(m) = \begin{cases} \log \tilde{K}, & m = 0 \\ -\tilde{a}(m) - \sum_{k=1}^{m-1} \frac{k}{m}\tilde{c}(k)\,\tilde{a}(m-k), & 1 \le m \le N \end{cases} \qquad (3)$$

where $a(0) = 1$.

$a(i)$ は線形予測係数、α は周波数圧伸係数、K はゲイン項を示す。ここで、$\alpha = 0$ とすると LPC ケプストラムの計算式となり、サンプル周波数 8kHz で $\alpha = 0.31$、サンプル周波数 12kHz で $\alpha = 0.37$、サンプル周波数 16kHz で $\alpha = 0.42$ とするとメル尺度を良く近似する LPC メルケプストラムの計算式となる。

3.2 MFCC(Mel-Frequency Cepstrum Coefficient)

N 次の MFCC $b(n)$ は、

$$b(n) = \sqrt{\frac{2}{M}} \sum_{m=1}^{M} S(m) \cos\left(\frac{\pi n}{M}(m-0.5)\right), \quad 1 \le n \le N \qquad (4)$$

により求められる [1]。ここで、$S(m)$ はメルスケール $Mel(f)$ を M 個に等分割したときの m 番目の対数フィルタバンク振幅である。また、周波数 f[Hz] に対するメルスケールは、

$$Mel(f) = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \qquad (5)$$

によって与えられる。

3.3 実験内容

上述した、LPC(メル) ケプストラム、MFCC の認識性能を比較するためにベースライン実験と同じ学習、評価データを用い音素認識実験を行った。サンプル周波数を 12kHz、16kHz とし、サンプル周波数と音響ベクトルを除く表 1に示す条件で分析を行った LPC ケプストラム、LPC メルケプストラム、MFCC を特徴パラメータ (ケプストラム次数:8,12,16) とした。これらの特徴パラメータとパワー、及びそれぞれの 1 次回帰係数の合計 2(N+1) (N: ケプストラム次数) 次元を特徴ベクトルとし、ベースライン実験同様に音素認識実験を行った。

ここで、LPC 分析に用いた分析次数はケプストラム次数に関わらず、サンプル周波数 12kHz の場合 16 次、サンプル周波数 16kHz の場合 20 次とした。また、MFCC 計算に用いたフィルタバンク次数はサンプル周波数に関わらず、N+4 とした。それぞれの特徴ベクトルに対する音素認識率を図 1、2、表 3、4に示す。

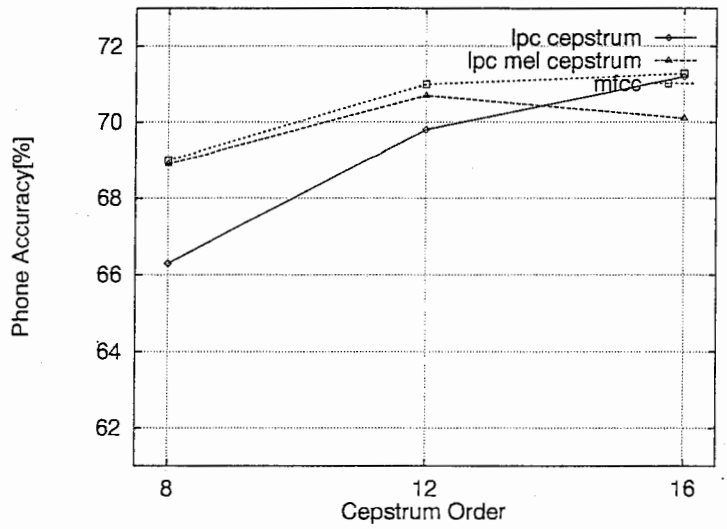

図 1: LPC(メル) ケプストラム、 MFCC の比較 (サンプル周波数 12kHz)

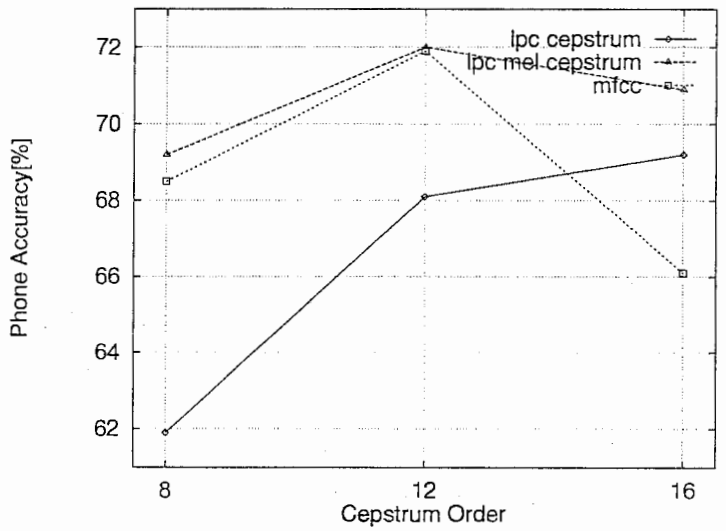

図 2: LPC(メル) ケプストラム、 MFCC の比較 (サンプル周波数 16kHz)

表 3: 比較実験の音素認識率 (%)(Sampling rate 12kHz)

ケプストラム次数	LPC ケプストラム (男性 / 女性 / 全体)	LPC メルケプストラム (男性 / 女性 / 全体)	MFCC (男性 / 女性 / 全体)
8	63.5/68.0/66.3	66.1/70.6/68.9	66.5/70.5/69.0
12	66.2/72.0/69.8	67.9/72.5/70.7	68.9/72.2/71.0
16	66.6/74.0/71.2	66.3/72.4/70.1	69.0/72.7/71.3

表 4: 比較実験の音素認識率 (%)(Sampling rate 16kHz)

ケプストラム次数	LPC ケプストラム (男性 / 女性 / 全体)	LPC メルケプストラム (男性 / 女性 / 全体)	MFCC (男性 / 女性 / 全体)
8	58.1/64.2/61.9	64.6/71.8/69.2	65.0/70.6/68.5
12	63.5/70.6/68.1	68.2/74.2/72.0	68.3/74.0/71.9
16	65.0/71.7/69.2	67.5/72.9/70.9	63.5/67.8/66.1

3.4　実験結果、考察

図 1、2より、LPC ケプストラムは、ケプストラムの次元減少に伴い認識率が低下している。これは、LPC ケプストラムが周波数領域において、線形であるため、認識に重要と考えられる低周波数帯が次元の減少とともに、分析が行いにくくなっていると予想される。逆に、メルスケールを用いている LPC メルケプストラム、MFCC はケプストラム次数が減少しても認識率の低下が少ない。

以上の結果から、今回実験に用いた自然発話音声データベースに対しては、サンプル周波数 12kHz においては、MFCC が、サンプル周波数 16kHz においては、LPC メルケプストラムがケプストラム次数の変動に対しても安定した認識性能を示すため、特徴パラメータとして有効であると考えられる。有効と考えられたサンプル周波数 12kHz、MFCC12 次元、サンプル周波数 16kHz、LPC メルケプストラム 12 次元の話者ごとの認識結果を付録 E.1、E.2に示す。

4 ポーズモデルに関する実験

4.1 実験内容

ベースライン実験の音素認識結果より、無音部分への挿入誤りが数多く見受けられる (付録 D.1参照)。これは、ベースラインで使用したポーズモデルが 1 状態で表されているため、長い無音に対応できていないことが原因であると考えられる。そこで、無音部分を表すポーズモデルの学習方法、状態数を変更し認識性能を比較した。

ポーズモデルはラベル学習後、有音モデルと共に発話の前後 30msec に無音を追加したものを用い連結学習を行っていた。しかし、この方法では、(1) 有音部分付近の無音データしか学習に用いていない、(2) 30msec と非常に短くかつ固定長の無音データに対して学習を行っていることが問題であると考えられる。このため、無音モデルはラベル学習により得られるものを用い、さらに従来 1 状態で作成していた無音モデルを複数状態に変えて認識性能を比較した。

実験は、ベースライン実験と同様に表 1に示す分析条件で行った。ポーズモデルの状態数の変更は、状態数を変更したポーズモデルを新たに作成し、ベースライン実験の音響モデルのポーズモデルと入れ替え新たな音響モデルとし、認識実験を行った。

4.2 実験結果、考察

ポーズモデルを変化させた音素認識実験結果を図 3、表 5に示す。

図 3より、ポーズモデルを含む連結学習を行うことにより認識性能が低下していることが確認できる。これは、ポーズモデルが 30mse という短い時間の連結学習を行うため、長い無音に対応ができなくなり、挿入誤りを引き起こしてしまうと考えられる。

また、ラベル学習のみでポーズモデルの状態数を変化させたときの認識にした全音素数に対する各誤りの割合を図 4、表 6に示す。これより、挿入誤りの減少には状態数増加が有効であることが分かる。しかし、状態数増加と共に削除誤りの増加が起こってしまう。そのため、連結学習を行わない状態数 5 のポーズモデルが最も有効であると推測される。有効であると推測された連結学習を行わない状態数 5 のポーズモデルを用いた話者ごとの結果を付録 F.1に示す。

表 5: ポーズモデル変化による音素認識率 (%)

状態数	男性	女性	全体
1(連結学習有)	66.60	74.04	71.21
1	67.08	74.55	71.71
3	68.97	76.88	73.87
5	69.24	77.14	74.14
7	69.23	76.84	73.95
9	67.84	74.61	72.04

表 6: ポーズモデル変化による誤りの割合 (%)(全音素数 21014 個)

状態数	挿入誤り	削除誤り	置換誤り
1(連結学習有)	9.5	5.5	13.8
1	9.2	5.5	13.6
3	6.9	5.8	13.4
5	6.5	5.8	13.5
7	6.2	6.4	13.5
9	5.7	8.8	13.4

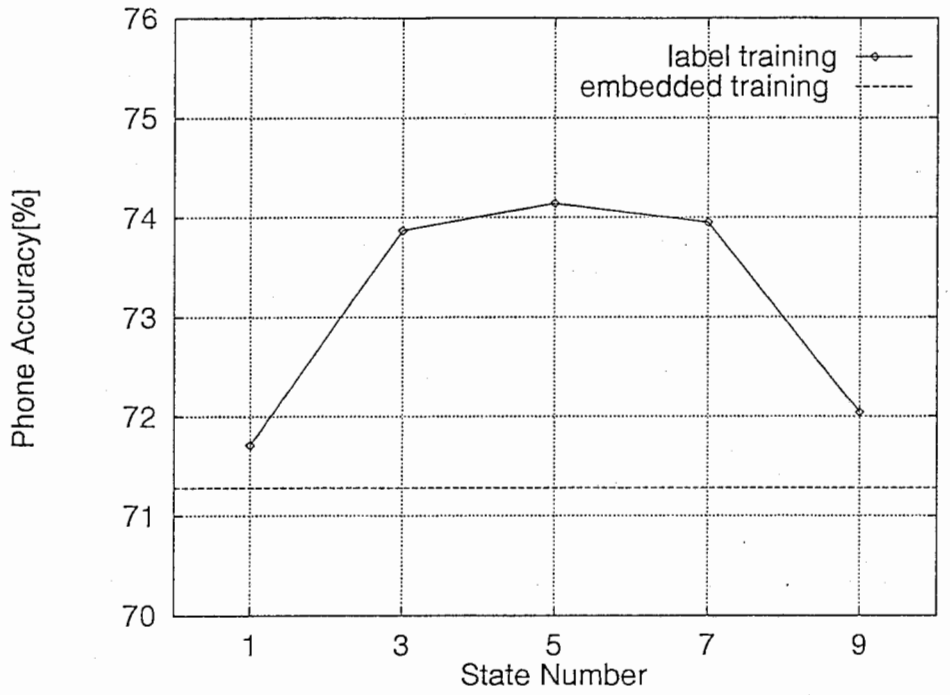

図 3: ポーズモデル変更による音素認識率

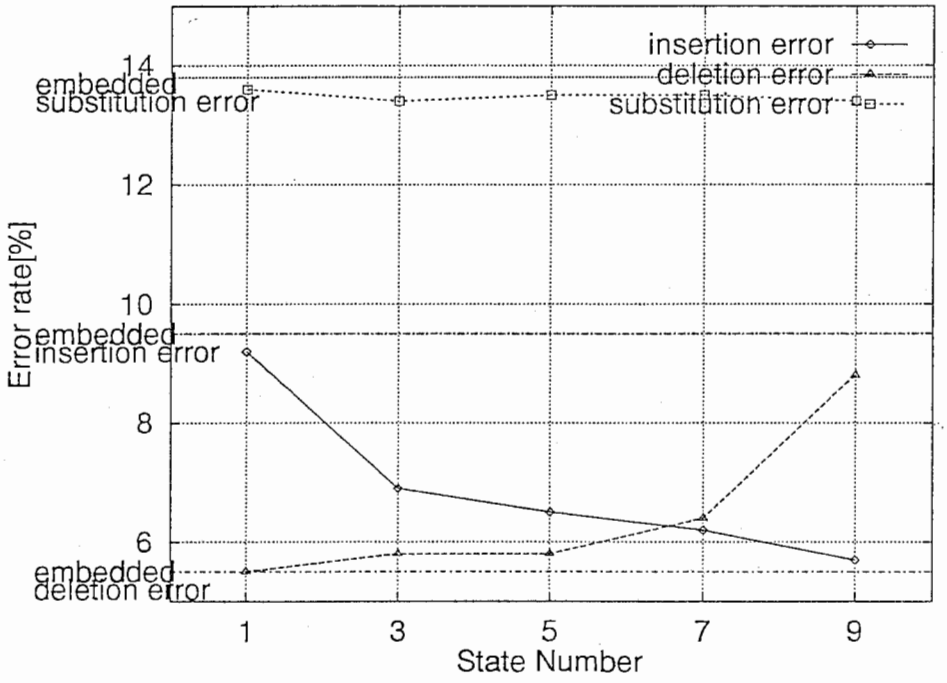

図 4: ポーズモデルの状態数と誤り割合

5 むすび

本報告では、LPC(メル) ケプストラム、MFCC に対する認識性能の比較実験を行い、サンプル周波数 12kHz においては MFCC が、サンプル周波数 16kHz においては LPC メルケプストラムが次元の変動に対し安定した認識性能を得ることが可能であることを示した。

また、サンプル周波数 12kHz、LPC ケプストラム 16 次におけるポーズモデルとして、連結学習を行わない、ラベル学習のみの 5 状態のポーズモデルが高い認識性能を得ることを示した。本報告では、他の条件下でのポーズモデルの比較を行わなかったが、他の条件下でも本報告の結果に類似すると推測される。そのためポーズモデルは連結学習を行わない複数の状態が有効であると考えられる。

参考文献

[1] Cambridge University Engineering Department Speech Group. HTK: Hidden Markov Model Toolkit V1.5. manual, September 1993.

[2] A. Nakamura, S. Matsunaga, T. Shimizu, M. Tonomura, and Y. Sagisaka. Japanese Speech Databases for Robust Speech Recognition. In *ICSLP*, pp. 2199–2202, Philadelphia, 1996.

[3] M. Ostendorf and H. Singer. HMM Topology Design Using Maximum Likelihood Successive State Splitting. *Computer Speech and Language*, Vol. 11, No. 1, pp. 17–41, 1997.

[4] H. Singer, M.Tonomura, Q. Huo, J. Ishii, T. Fukada, and M. Schuster. Baseline Acoustic Models for the Spoken Language Database(SDB/SLDB). Technical Report TR-IT-0206, ATR, 1997.

[5] 徳田恵一, 小林隆夫, 今井聖. メル一般化ケプストラムの再帰的計算法. 信学論 (A), Vol. J71-A, No. 1, pp. 128–131, January 1988.

[6] 嵯峨山茂樹. 音声認識のための音声分析とラベル変換. Technical Report TR-I-0347, ATR, 1993.

付録 A パワー項に関する実験

A.1 実験内容

音声信号の音の大きさを表しているパワー項を基準実験の特徴ベクトルから除き、残りの 33 次元を特徴ベクトルとし実験を行った。パワー項を除く以外はベースライン実験と同様に実験を行った。パワー項を除くことにより、各話者の声の大小、マイクロフォンとの距離などの影響を受けにくくなると考えられる。

A.2 実験結果、考察

結果を表 7に示す。表 7により、クリーンなデータベースを用いた場合、パワー項は音素識別能力を持っていることが確認された。

表 7: 認識結果 (音素認識率 (%))

	男性	女性	全体
ベースライン	66.6	74.0	71.2
パワー項を除く	63.7	70.9	68.2

付録 B HMnet 作成条件に関する実験

B.1 実験内容

ベースライン実験と同様表 1に示した分析条件、学習データ、評価データを用い、HMnet 作成条件のみを変更し認識性能、HMnet 作成時間の比較を行った。表 8に比較を行った HMnet 作成条件を示す。

表 8: HMnet 作成条件

実験番号	初期状態数	初期経路長	最大経路長	当該音素間の共有
ベースライン	27	3	4	認める
Ex-1	75	3	4	認めない
Ex-2	27	3	3	認める
Ex-3	75	3	3	認めない
Ex-4	52	4	4	認める
Ex-5	100	4	4	認めない

B.2 実験結果、考察

表 9に各実験に対する認識結果、図 5、6、7に、HMnet topology が 1000 状態に至るまでの HMnet topology 作成時間を示す。

表 9より、初期状態数は他の作成条件が等しい場合には、認識性能にほぼ影響を与えないことが確認される。これは、800 状態まで分割が行われると当該音素間の共有がほぼ見られなくなるためと考えられる。また、各音素毎継続時間が異なるため一定の時間方向への分割では認識性能が低下していることがわかる。よって、HMnet 作成条件として有効なのは、時間方向に分割を可能にし、当該音素間の状態共有を行う、初期状態数 27 と考えられる。

また、図 5、6、7より、再学習時間は一定状態数 (200 状態) 以上となれば、初期状態数の影響を受けず各条件、一定時間で学習が行われることが分かる。状態分割時間に関しても、実験 2 を除く他の実験は同様の結果を示した。しかし、実験 2 に関しては状態分割時間の急激な増加がみられる。この急激な増加原因は不明である。

表 9: HMnet 作成条件変更実験の実験結果 (音素認識率 (%))

実験番号	男性	女性	全体
ベースライン	66.6	74.0	71.2
Ex-1	67.5	73.4	71.1
Ex-2	64.9	70.6	68.4
Ex-3	65.1	71.0	68.7
Ex-4	65.9	72.8	70.1
Ex-5	65.7	72.8	70.1

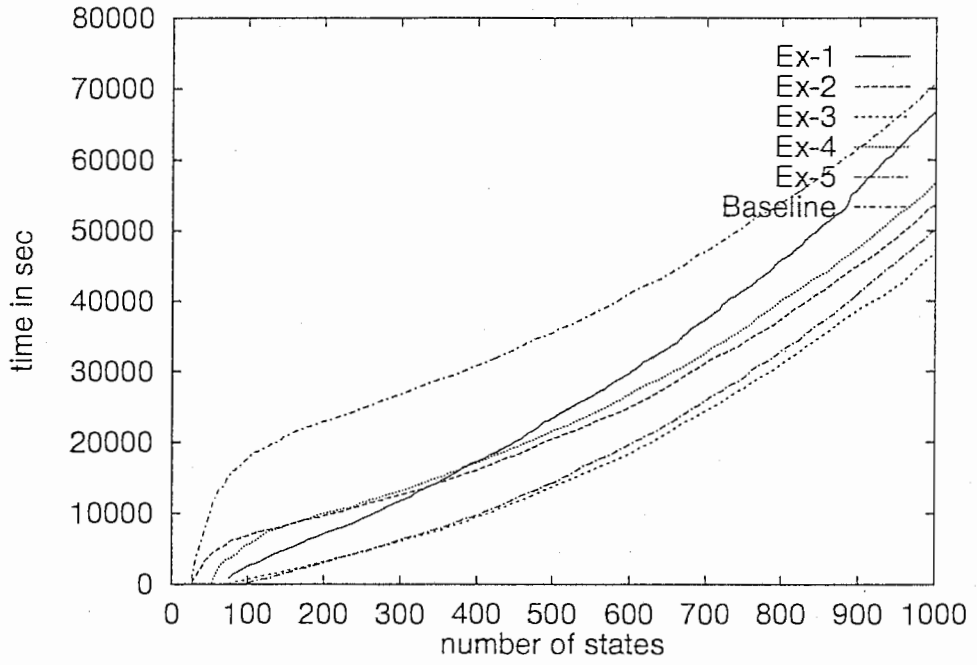

図 5: 初期状態変化による HMnet topology 作成時間 (全時間)

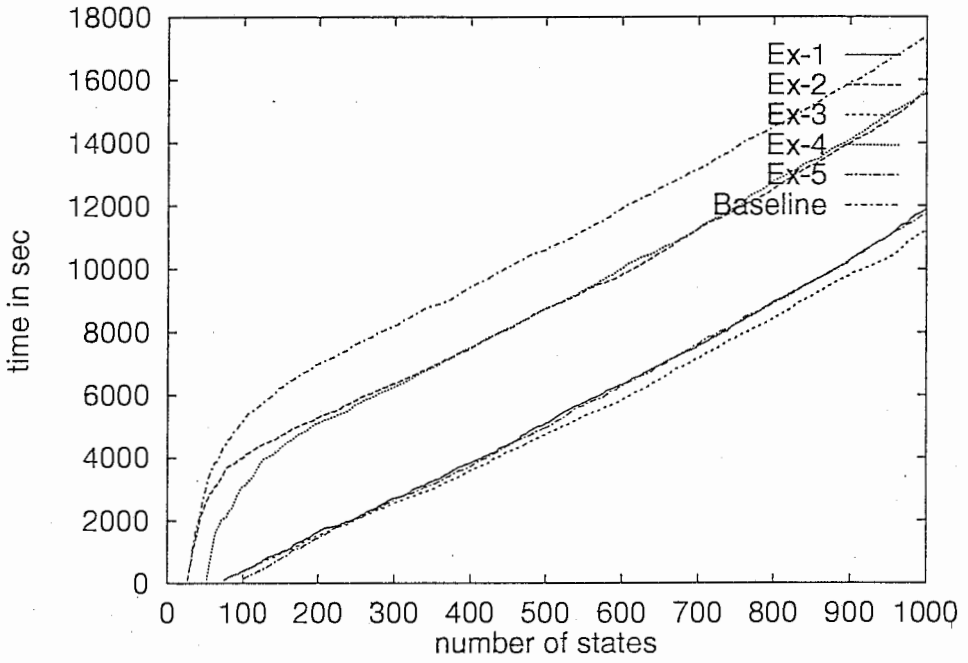

図 6: 初期状態変化による HMnet topology 作成時間 (再学習時間)

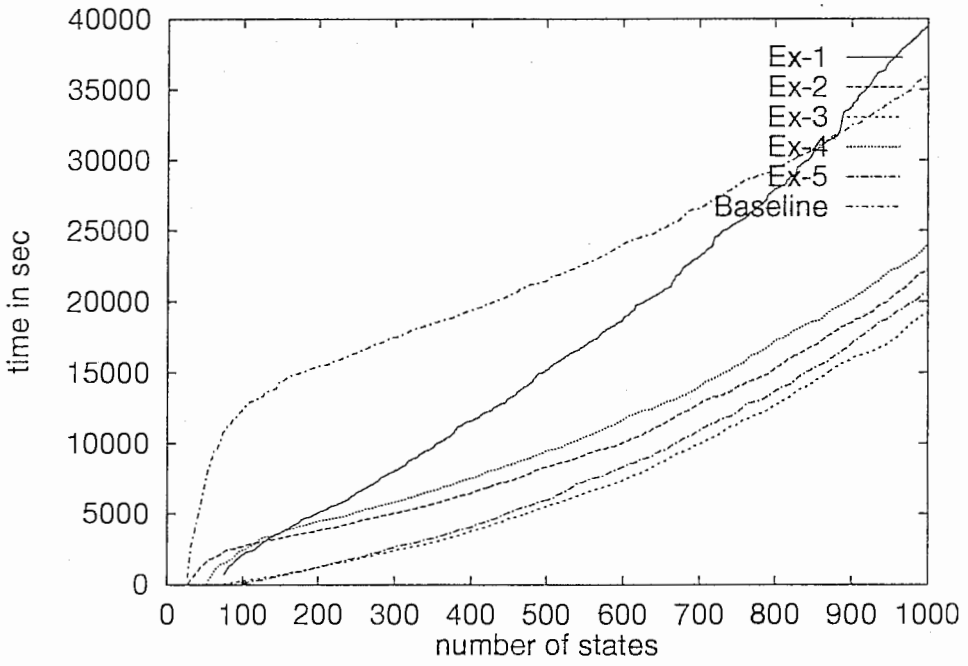

図 7: 初期状態変化による HMnet topology 作成時間 (状態分割時間)

表 10: HMnet 作成時学習状態変化実験の認識結果 (音素認識率 (%))

実験番号	男性	女性	全体
ベースライン	66.6	74.0	71.2
Ex-6	65.0	72.6	69.7

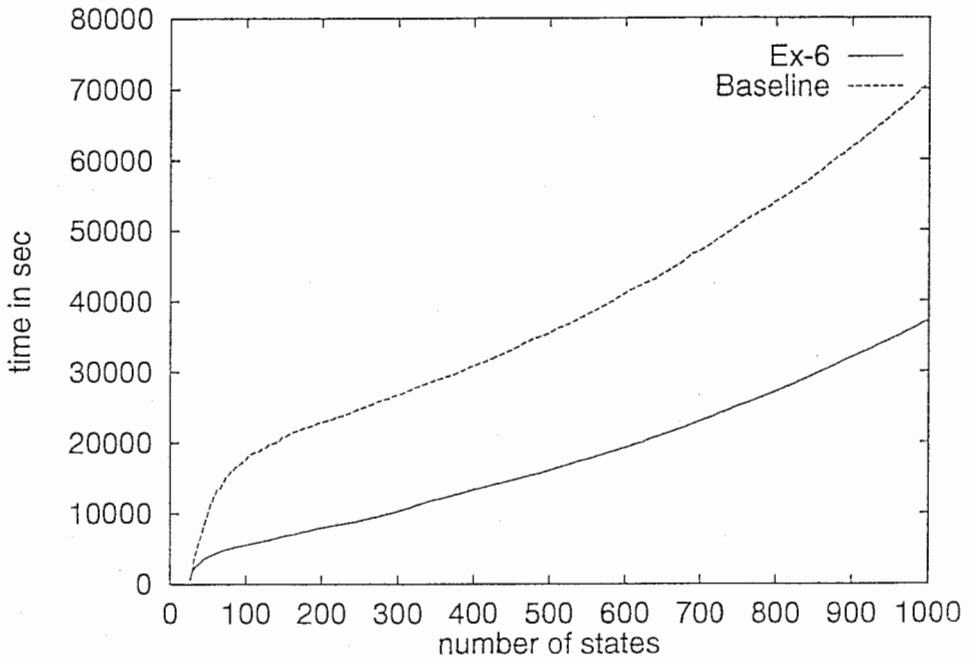

図 8: 再学習影響を受ける状態数変化による HMnet 作成時間 (全時間)

付録 C　HMnet 作成時学習状態の変化に関する実験

C.1　実験内容

ベースライン実験は、HMnet topology 作成時に状態分割を行った状態から影響を受ける全状態に対し再学習 (連結学習、分割情報学習) を行っていた。本実験 (Ex-6) においては、分割を行った状態から強く影響を受ける状態、すなわち分割を行った状態に接続する状態のみに限定を行い再学習を行った。これより、再学習を行う状態数の変化による認識性能、分割時間、再学習時間の比較を行った。

C.2　実験結果、考察

表 10に、本実験に対する認識結果、図 8、9、10に再学習を行う状態数の変化による各時間の変化を示す。表 10、8より、再学習を行なう状態数を減少させる、すなわち、少ない再学習状態数で近似することにより、HMnet topology 作成時間を約半分に減少することができる、しかし、認識性能は 1.5(%) 劣化してしまう。そのため、目的によって再学習を行なう状態数を変更することが望まれる。

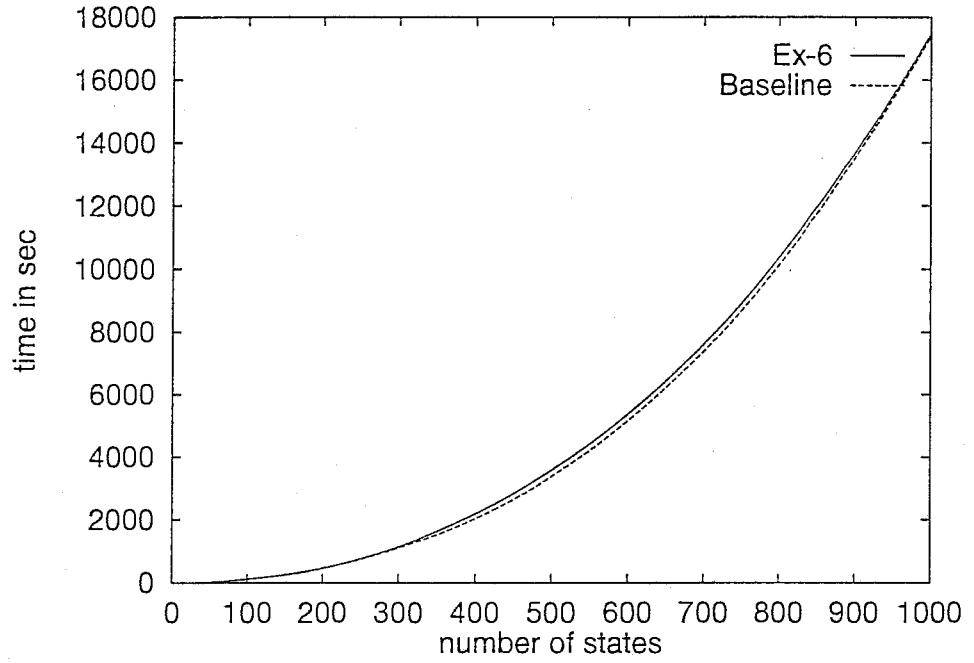

図 9: 再学習影響を受ける状態数変化による HMnet 作成時間 (状態分割, 再学習以外の時間)

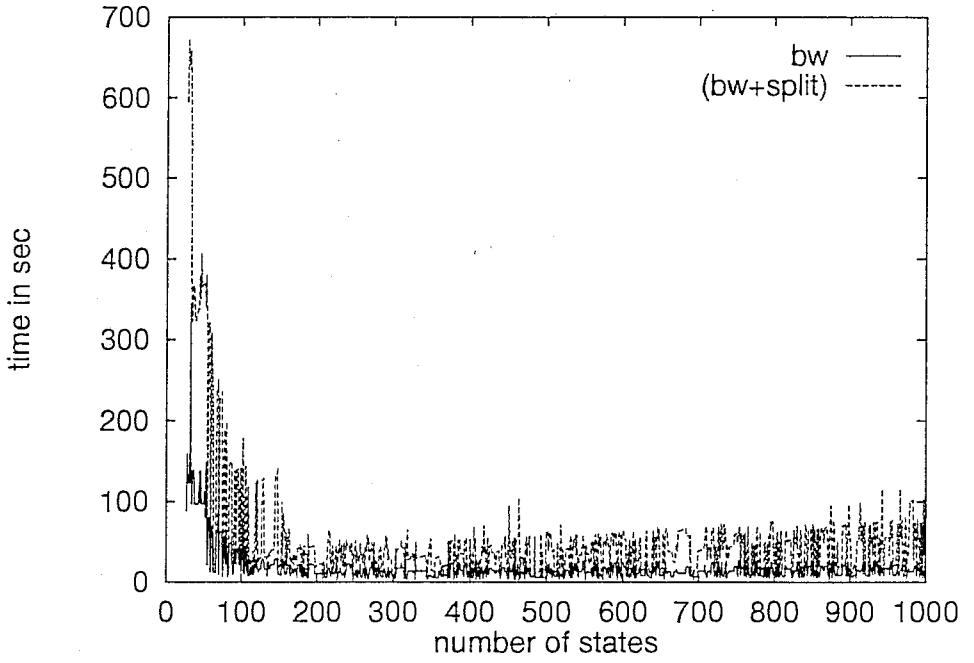

図 10: 各状態数による状態分割時間、再学習時間

付録 D　ベースライン実験の詳細

D.1　各話者毎の認識結果

話者	音素認識率	全音素数	挿入誤り数	削除誤り数	置換誤り数
TAC70015.A	66.23	453	45	16	92
TAC70016.A	64.94	348	16	38	68
TAC70017.A	78.95	247	6	10	36
TAC70019.A	75.64	390	32	14	49
TAC70021.A	55.94	429	57	20	112
TAC70022.A	57.96	509	31	88	95
TAC70023.A	65.57	485	85	15	67
TAC70101.A	86.94	444	25	8	25
TAC70102.A	75.29	526	68	8	54
TAC70103.A	65.34	277	19	24	53
TAC70201.A	81.04	501	38	13	44
TAC70202.A	69.82	570	22	64	86
TAC70203.A	68.44	526	85	12	69
TAC70301.A	71.14	447	60	16	53
TAC70303.A	87.09	457	28	13	18
TAC70304.A	41.38	261	64	6	83
TCC70103.A	75.96	366	32	15	41
TCC70109.A	70.48	376	33	23	55
TCC70201.A	65.98	338	24	42	49
TCC70212.A	56.33	616	96	29	144
TCC70307.A	73.24	568	39	29	84
TCC71001.B	68.40	864	68	71	134
TCC71007.A	80.06	617	60	13	50
TCC71008.A	72.37	684	48	49	92
TCC71016.A	76.27	573	78	2	56
TCC71035.A	75.07	373	39	17	37
TCS70004.B	66.10	1109	50	112	21
TCS70010.A	70.08	391	12	44	61
TCS70013.A	81.00	400	32	15	29
TCS70020.A	63.97	433	85	11	60
TCS70023.A	77.08	903	110	30	67
TCS70025.A	76.43	454	22	24	61
TCS70028.A	76.99	352	35	7	39
TCS70034.A	66.77	626	82	23	103
TCS70047.A	75.50	400	39	23	36
TCS70055.A	71.49	677	31	80	82
TCS70059.A	77.37	358	31	16	34
TCS70070.A	65.33	300	38	24	42
TCS70074.A	77.30	304	23	15	31
TCS70082.A	71.74	644	68	24	90
TSC71005.B	75.00	532	45	12	76
TSC71013.A	70.32	886	102	37	124
全体	71.21	21014	2003	1152	2895

D.2 ベースラインの設定

- 特徴パラメータ作成のコンフィグレーション

```
# sample options list : ATRwave96 : I/Ocontrol
I/Ocontrol:inputFormat=NoHeader
I/Ocontrol:inputParamSize=120
I/Ocontrol:inputParamType=short
I/Ocontrol:inputFd=stdin
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:outputFormat=NoHeader
I/Ocontrol:outputParamSize=34
I/Ocontrol:outputParamType=float
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputByteorder=BigEndian

ATRwave2cep:inputParameter=waveRaw
ATRwave2cep:Preemphasis=0.98
ATRwave2cep:FrameLength=20
ATRwave2cep:FrameShift=10
ATRwave2cep:SamplingFrequency=12000
ATRwave2cep:TimeWindow=hamming
ATRwave2cep:LagWindowFactor=0.01
ATRwave2cep:LpcOrder=16
ATRwave2cep:CepstrumOrder=16
ATRwave2cep:FrequencyWarping=linear
ATRwave2cep:AnalysisType=lpc
ATRwave2cep:DebuggingLevel=0

ATRcep2para:CepstrumOrder=16
ATRcep2para:LDA=
ATRcep2para:DeltaCepstrumWindow=9
ATRcep2para:deltaCepstrumPadding=zero
ATRcep2para:DDCepstrumWindow=9
ATRcep2para:DDCepstrumPadding=zero
ATRcep2para:rho=1.0
ATRcep2para:OutputParameter=pow+cep(16)+dpow+dcep(16)

# sample options list : ATRexpandSample :
ATRexpand:samplingFrequency=12000
ATRexpand:frameShift=10
ATRexpand:outputParamSize=34
ATRexpand:inputFd
ATRexpand:outputFd
ATRexpand:debuggingLevel=ON
ATRexpand:htkFlag=0
ATRexpand:outputFormat=SSSData
ATRexpand:exec="/home/atra19/stsuge/tmp/atra52.6052.4 -config=/home/atra19/stsuge/tmp/atra52.6052.5"
```

- 認識のためのコンフィグレーション

```
#I/Ocontrol config : Mon May 26 13:53:32 1997
I/Ocontrol:rpcTable=
I/Ocontrol:rpcNumber=3
I/Ocontrol:outputByteorder=BigEndian
I/Ocontrol:outputFd=stdout
I/Ocontrol:outputParamType=
I/Ocontrol:outputParamSize=
I/Ocontrol:outputFormat=Lattice
I/Ocontrol:inputByteorder=BigEndian
I/Ocontrol:inputFd=/data/atra55/itlusers/stsuge/RPARASDB/TSC71013.A.FSYNC
I/Ocontrol:inputParamType=float
I/Ocontrol:inputParamSize=34
I/Ocontrol:inputFormat=FrameSync

#ATRresult config : Mon May 26 13:53:32 1997
ATRresult:answer=/dept1/work1/V1/data/TSC71013.A.TRS
ATRresult:dp_weight=1.0,1.0,1.0
ATRresult:pause_symbol=-
ATRresult:UTT_END=6
ATRresult:UTT_START=5
ATRresult:re_beam=
ATRresult:N_best=1
ATRresult:N_best_out=stdout
ATRresult:lattice_out=stdout
#ATRlattice config file : Mon May 26 13:53:32 JST 1997
ATRlattice:lexicon=/dept1/work1/V1/model/LEX.P
ATRlattice:amname=./ModelMLSSS/A230/HMnet_filled.emb.5.800
ATRlattice:active_model=1
ATRlattice:lmscale=4.000000,8.000000
ATRlattice:wdpenalty=0,0
ATRlattice:ngram=Class-2,/dept1/work1/V1/model/LM.P
ATRlattice:beam=30,30
ATRlattice:work_area=200,50
ATRlattice:frame_shift=10
ATRlattice:pause_symbol=-
ATRlattice:dimension=34
ATRlattice:max_allophone=5000
ATRlattice:phone_boundary=OFF
ATRlattice:word_merge=non
ATRlattice:UTT_START=5
ATRlattice:UTT_END=6
ATRlattice:backward_frame=-1
```

付録 E 比較実験の詳細

E.1 サンプル周波数 12kHz、MFCC12 次元の各話者毎の認識結果

話者	音素認識率	全音素数	挿入誤り数	削除誤り数	置換誤り数
TAC70015.A	67.99	453	49	17	79
TAC70016.A	67.24	348	14	37	63
TAC70017.A	87.85	247	2	7	21
TAC70019.A	80.00	390	29	15	34
TAC70021.A	62.00	429	42	26	95
TAC70022.A	70.53	509	32	30	88
TAC70023.A	67.63	485	78	16	63
TAC70101.A	85.81	444	21	9	33
TAC70102.A	81.75	526	50	4	42
TAC70103.A	67.87	277	13	22	54
TAC70201.A	79.04	501	37	13	55
TAC70202.A	70.70	570	16	71	80
TAC70203.A	66.35	526	62	25	90
TAC70301.A	73.15	447	44	14	62
TAC70303.A	85.12	457	25	12	31
TAC70304.A	46.36	261	53	10	77
TCC70103.A	74.32	366	27	13	54
TCC70109.A	71.81	376	38	26	42
TCC70201.A	65.38	338	23	42	52
TCC70212.A	60.23	616	72	38	135
TCC70307.A	73.24	568	49	38	65
TCC71001.B	73.38	864	39	67	124
TCC71007.A	74.55	617	63	17	77
TCC71008.A	66.67	684	52	60	116
TCC71016.A	68.41	573	76	13	92
TCC71035.A	78.55	373	33	11	36
TCS70004.B	55.28	1109	47	138	311
TCS70010.A	62.15	391	18	56	74
TCS70013.A	76.50	400	39	14	41
TCS70020.A	61.66	433	74	19	73
TCS70023.A	74.64	903	90	32	107
TCS70025.A	80.40	454	14	24	51
TCS70028.A	81.82	352	24	6	34
TCS70034.A	70.93	626	82	28	72
TCS70047.A	79.50	400	34	17	31
TCS70055.A	68.83	677	33	90	88
TCS70059.A	70.39	358	23	44	39
TCS70070.A	68.00	300	27	24	45
TCS70074.A	80.92	304	11	19	28
TCS70082.A	74.22	644	61	25	80
TSC71005.B	68.98	532	42	23	100
TSC71013.A	66.03	886	105	49	147
全体	70.95	21014	1763	1261	3081

E.2　サンプル周波数 16kHz、LPC メルケプストラム 12 次元の各話者毎の認識結果

話者	音素認識率	全音素数	挿入誤り数	削除誤り数	置換誤り数
TAC70015.A	70.86	453	53	20	59
TAC70016.A	60.06	348	29	36	74
TAC70017.A	82.59	247	10	6	27
TAC70019.A	80.00	390	23	11	44
TAC70021.A	64.34	429	45	19	89
TAC70022.A	60.12	509	47	46	110
TAC70023.A	68.04	485	80	19	56
TAC70101.A	88.51	444	24	8	19
TAC70102.A	77.00	526	66	10	45
TAC70103.A	70.76	277	10	23	48
TAC70201.A	82.24	501	33	11	45
TAC70202.A	75.96	570	18	58	61
TAC70203.A	73.19	526	58	7	76
TAC70301.A	72.04	447	53	13	59
TAC70303.A	90.37	457	20	10	14
TAC70304.A	49.04	261	63	11	59
TCC70103.A	68.58	366	37	39	39
TCC70109.A	71.01	376	26	24	59
TCC70201.A	68.64	338	26	39	41
TCC70212.A	58.44	616	85	42	129
TCC70307.A	73.42	568	50	32	69
TCC71001.B	70.49	864	72	72	111
TCC71007.A	78.12	617	65	10	60
TCC71008.A	68.57	684	55	54	106
TCC71016.A	72.60	573	94	9	54
TCC71035.A	77.48	373	46	12	26
TCS70004.B	62.76	1109	47	166	200
TCS70010.A	63.68	391	22	52	68
TCS70013.A	80.50	400	37	12	29
TCS70020.A	62.36	433	74	13	76
TCS70023.A	74.42	903	122	29	80
TCS70025.A	78.19	454	17	25	57
TCS70028.A	86.36	352	19	5	24
TCS70034.A	73.00	626	70	19	80
TCS70047.A	77.75	400	34	15	40
TCS70055.A	70.01	677	40	93	70
TCS70059.A	80.45	358	21	15	34
TCS70070.A	65.00	300	29	22	54
TCS70074.A	79.28	304	24	11	28
TCS70082.A	73.76	644	67	23	79
TSC71005.B	75.94	532	48	13	67
TSC71013.A	65.46	886	117	46	143
全体	72.00	21014	1976	1200	2708

付録 F　ポーズモデルに関する実験の詳細

F.1　ラベル学習、5状態のポーズモデルの各話者毎の認識結果

話者	音素認識率	全音素数	挿入誤り数	削除誤り数	置換誤り数
TAC70015.A	67.55	453	44	17	86
TAC70016.A	66.09	348	9	38	71
TAC70017.A	80.97	247	4	10	33
TAC70019.A	77.18	390	15	18	56
TAC70021.A	61.31	429	36	19	111
TAC70022.A	60.71	509	22	91	87
TAC70023.A	74.23	485	36	21	68
TAC70101.A	88.06	444	15	9	29
TAC70102.A	77.38	526	58	10	51
TAC70103.A	67.51	277	14	23	53
TAC70201.A	82.44	501	28	17	43
TAC70202.A	70.53	570	16	68	84
TAC70203.A	77.00	526	49	15	57
TAC70301.A	76.06	447	39	17	51
TAC70303.A	89.50	457	16	15	17
TAC70304.A	46.74	261	49	5	85
TCC70103.A	78.69	366	18	20	40
TCC70109.A	74.20	376	24	27	46
TCC70201.A	68.05	338	21	42	45
TCC70212.A	59.90	616	73	32	142
TCC70307.A	73.94	568	38	29	81
TCC71001.B	71.06	864	45	70	135
TCC71007.A	83.79	617	36	15	49
TCC71008.A	71.35	684	40	69	87
TCC71016.A	79.58	573	61	2	54
TCC71035.A	78.82	373	24	18	37
TCS70004.B	68.80	1109	26	112	208
TCS70010.A	70.33	391	13	44	59
TCS70013.A	81.25	400	28	16	31
TCS70020.A	70.67	433	52	10	65
TCS70023.A	80.95	903	74	29	69
TCS70025.A	79.30	454	11	22	61
TCS70028.A	81.53	352	20	12	33
TCS70034.A	69.17	626	64	23	106
TCS70047.A	77.25	400	28	27	36
TCS70055.A	73.12	677	23	79	80
TCS70059.A	83.52	358	13	16	30
TCS70070.A	68.67	300	24	26	44
TCS70074.A	80.26	304	16	15	29
TCS70082.A	77.95	644	26	24	92
TSC71005.B	76.50	532	37	13	75
TSC71013.A	71.90	886	77	44	128
全体	74.14	21014	1362	1229	2844

付録 G　音響分析に関する実験の詳細

G.1　各実験を行ったバージョン

この実験バージョンは、比較実験においては認識結果から出力された *.err ファイル内 @MESSAGE(RPCLIB 04r04) を抜粋し決定した、ポーズモデル実験においては、ログファイル内の認識結果から同様の所を抜粋し、決定を行った。

表 11: サンプル周波数 12kHz 比較実験バージョン

ケプストラム次数	LPC ケプストラム	LPC メルケプストラム	MFCC
8	r04r02	r04r02	r04r02
12	r04r02	r04r02	r04r02
16	r04r01	r04r02	r04r02

表 12: サンプル周波数 16kHz 比較実験バージョン

ケプストラム次数	LPC ケプストラム	LPC メルケプストラム	MFCC
8	r04r03	r04r04	r04r04
12	r04r03	r04r04	r04r04
16	r04r03	r04r03	r04r03

表 13: ポーズモデル変化実験バージョン

状態数	実験バージョン
1(連結学習有)	r04r01
1	r04r02
3	r04r02
5	r04r02
7	r04r02
9	r04r02